

第2章 关系模型



沈明玉

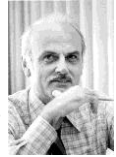
合肥工业大学
Hefei University of Technology

计算机与信息学院

关系数据库之父

❖ 美国 IBM 公司的 E.F.Codd

- ☞ 埃德加·弗兰克·科德 (Edgar Frank Codd, 1923-2003)
- ☞ 1970年提出关系数据模型
“A Relational Model of Data for Large Shared Data Banks”,
《Communication of the ACM》, 1970
- ☞ 之后, 提出了关系代数和关系演算的概念
- ☞ 1972年提出了关系的第一、第二、第三范式
- ☞ 1974年提出了关系的BC范式



合肥工业大学
Hefei University of Technology

计算机与信息学院

第2章 关系模型

主要内容:

- 2.1 关系模型的数据结构及形式化定义
- 2.2 关系模型的数据操作
- 2.3 关系模型的完整性约束
- 2.4 关系代数语言

合肥工业大学
Hefei University of Technology

计算机与信息学院

第2章 关系模型

2.1 关系模型的数据结构及形式化定义

2.1.1 关系

● 单一的数据结构-关系

现实世界的**实体**以及实体间的各种联系均用关系来表示

● 逻辑结构-二维表

从用户角度, 关系模型中数据的逻辑结构是一张二维表

● 建立在集合代数的基础上

合肥工业大学
Hefei University of Technology

计算机与信息学院

2.1 数据结构及形式化定义

● 相关定义

(1) 域 (Domain): 一组具有相同数据类型的值的集合。

例如: 整数、实数、介于某个取值范围的整数、指定长度的字符串集合、{‘男’, ‘女’}等。

(2) 笛卡尔积 (Cartesian Product) - 基于域的一种集合运算

给定一组域 D_1, D_2, \dots, D_n , 这些域中可以有相同的。

D_1, D_2, \dots, D_n 的笛卡尔积为:

$$D_1 \times D_2 \times \dots \times D_n = \{(d_1, d_2, \dots, d_n) \mid d_i \in D_i, i=1, 2, \dots, n\}$$

合肥工业大学
Hefei University of Technology

计算机与信息学院

2.1 数据结构及形式化定义

✓ 所有域的所有取值的一个组合

✓ 不能重复

✓ 笛卡尔积中每一个元素 (d_1, d_2, \dots, d_n) 叫作一个 n 元组或简称**元组**。

✓ 笛卡尔积元素 (d_1, d_2, \dots, d_n) 中的每一个值 d_i 叫作一个**分量**。

✓ **基数**: 若 $D_i (i=1, 2, \dots, n)$ 为有限集, 其基数为 $m_i (i=1, \dots, n)$, 则 $D_1 \times D_2 \times \dots \times D_n$ 的基数 M 为:

$$M = \prod_{i=1}^n m_i$$

合肥工业大学
Hefei University of Technology

计算机与信息学院

2.1 数据结构及形式化定义

设有3个域:

导师 = {张清玫, 刘逸}

专业 = {计算机专业, 信息专业}

研究生 = {李勇, 刘晨, 王敏}

笛卡尔积 = 导师 \times 专业 \times 研究生

有意义吗?

导师	专业	研究生
张清玫	计算机专业	李勇
张清玫	计算机专业	刘晨
张清玫	计算机专业	王敏
张清玫	信息专业	李勇
张清玫	信息专业	刘晨
张清玫	信息专业	王敏
刘逸	计算机专业	李勇
刘逸	计算机专业	刘晨
刘逸	计算机专业	王敏
刘逸	信息专业	李勇
刘逸	信息专业	刘晨
刘逸	信息专业	王敏

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

(3) 关系 (Relation)

$D_1 \times D_2 \times \dots \times D_n$ 的子集叫作在域 D_1, D_2, \dots, D_n 上的关系, 表示为:

$R(D_1, D_2, \dots, D_n)$

R : 关系名

n : 关系的目或度 (Degree)

关系也是一个二维表, 表的每行对应一个元组, 表的每列对应一个域。

导师	专业	研究生
张清玫	计算机专业	李勇
张清玫	计算机专业	刘晨
刘逸	信息专业	王敏

有意义的
数据!

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✓ 属性

- 关系系统中每一列必须起一个名字, 称为属性 (属性满足交换律!)

✓ 码

◆ 候选码 (Candidate key)

若关系中的某一属性组的值能唯一地标识一个元组, 且没有多余的属性, 则称该属性组为候选码。 (解释: 超码!)

◆ 主码 (Primary key)

若一个关系有多个候选码, 则选定其中一个为主码。

◆ 全码 (All-key)

所有属性组共同组成的候选码, 称为全码 (All-key)。

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✓ 主属性与非主属性

包含在任一候选码中的属性均为主属性, 否则为非主属性或非码属性。

✓ 三类关系

- 基本关系 (基本表): 实际存在的表, 是实际存储数据的逻辑表示。
- 查询表: 查询结果对应的表。
- 视图表: 由基本表或其他视图表导出的表, 是虚表。

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✓ 基本关系的性质

- ① 列是同质的;
- ② 不同的列可出自同一个域;
- ③ 列的顺序无所谓, 即列的次序可以任意交换;
- ④ 任意两个元组的候选码不能取相同的值;
- ⑤ 行的顺序无所谓, 即行的次序可以任意交换;
- ⑥ 分量必须取原子值, 即每一个分量都必须是不可分的数据项。

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

2.1.2 关系模式

关系模式是对关系的描述 (关系模式是型, 关系是值)。包括:

■ 元组集合的结构:

属性构成;

属性来自的域;

属性与域之间的映象关系。

■ 元组语义以及完整性约束条件

■ 属性间的数据依赖关系集合

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✓ 关系模式的形式化表示:

关系模式可以形式化地表示为:

$R(U, D, DOM, F)$

R 关系名

U 组成该关系的属性名集合

D 属性组 U 中属性所来自的域

DOM 属性向域的映像集合

F 属性间的数据依赖关系集合

合肥工业大学

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✓ 关系模式的简化形式:

$R(U)$ 或 $R(A_1, A_2, \dots, A_n)$

$R(U, F)$

■ R : 关系名

■ A_1, A_2, \dots, A_n : 属性名

注: 域名及属性向域的映像常常直接说明为属性的类型、长度。

合肥工业大学

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

✧ 关系模式与关系

✓ 关系模式 (型)

■ 对关系的描述, 静态的、稳定的。

✓ 关系 (值)

■ 关系模式在某一时刻的状态或内容;

■ 动态的、随时间不断变化的。

✓ 关系模式和关系往往统称为关系, 通过上下文加以区别

合肥工业大学

Hefei University of Technology 计算机与信息学院

2.1 数据结构及形式化定义

2.1.3 关系数据库

■ 在一个给定的应用领域中, 所有关系的集合构成一个关系数据库。

■ 理解: 关系数据库的型与值。

合肥工业大学

Hefei University of Technology 计算机与信息学院

第2章 关系模型

2.2 关系模型的数据操作

✧ 关系操作的特点

✓ 集合操作方式: 操作的对象和结果都是集合, 一次一集合的方式。

✧ 常用的关系操作

✓ 查询: 选择、投影、连接、除、并、交、差、笛卡尔积

✓ 数据更新: 插入、删除、修改

✓ 选择、投影、并、差、笛卡尔积是5种基本操作

合肥工业大学

Hefei University of Technology 计算机与信息学院

2.2 关系模型的数据操作

✧ 关系操作语言的分类

关系代数语言

◆ 用对关系的运算来表达查询要求。如: ISBL

关系演算语言: 用谓词来表达查询要求

◆ 元组关系演算语言: 谓词变元的基本对象是元组变量。如: ALPHA, QUEL

◆ 域关系演算语言: 谓词变元的基本对象是域变量。如: QBE

具有关系代数和关系演算双重特点的语言
如: SQL (Structured Query Language)

合肥工业大学

Hefei University of Technology 计算机与信息学院

第2章 关系模型

2.3 关系模型的完整性约束

关系的三类完整性

- ✓ 实体完整性
- ✓ 参照完整性
- ✓ 用户定义的完整性

合肥工业大学

计算机与信息学院



2.3 关系模型的完整性约束

❖ 实体完整性

若属性A是基本关系R的主属性，则属性A取值必须唯一且不能取空值。

❖ 参照完整性

若属性（或属性组）F是基本关系R的外码，它与基本关系S的主码K_s相对应（基本关系R和S不一定是不同的关系），则对于R中每个元组在F上的值必须为：

- 或者取空值（F的每个属性值均为空值）
- 或者等于S中某个元组的主码值

合肥工业大学

计算机与信息学院



2.3 关系模型的完整性约束

❖ 用户定义的完整性

- ✓ 针对某一具体关系数据库的约束条件，反映某一具体应用所涉及的数据必须满足的语义要求；
- ✓ 关系模型应提供定义和检验这类完整性的机制，以使用统一的系统的方法处理它们，而不要由应用程序承担这一功能。

合肥工业大学

计算机与信息学院



第2章 关系模型

2.4 关系代数语言

❖ 传统的集合运算

并(∪)、交(∩)、差(-)、笛卡尔积(×)

❖ 专门的关系运算

选择(σ)、投影(π)、连接(⋈)、除(÷)

❖ 运算符：算术运算符、逻辑运算符

合肥工业大学

计算机与信息学院



2.4 关系代数语言

2.4.1 传统的集合运算

(1) 并运算 (∪) :

$$R \cup S = \{t | t \in R \vee t \in S\}$$

问题：并操作对关系R和S的要求？

R ∪ S

A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁
a ₁	b ₃	c ₂

A	B	C
a ₁	b ₂	c ₂
a ₁	b ₃	c ₂
a ₂	b ₂	c ₁

合肥工业大学

计算机与信息学院



2.4 关系代数语言

(2) 交运算 (∩) :

$$R \cap S = \{t | t \in R \wedge t \in S\}$$

R ∩ S

A	B	C
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁

A	B	C
a ₁	b ₃	c ₂
a ₂	b ₂	c ₁

合肥工业大学

计算机与信息学院



2.4 关系代数语言

(3) 差运算 (-) :

$$R - S = \{t | t \in R \wedge t \notin S\}$$

A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁

A	B	C
a ₁	b ₂	c ₂
a ₁	b ₃	c ₂
a ₂	b ₂	c ₁

A	B	C
a ₁	b ₁	c ₁

合肥工业大学

计算机与信息学院



2.4 关系代数语言

(4) 广义笛卡尔积 (×) :

设: R: n目关系, k₁个元组
S: m目关系, k₂个元组

$$R \times S = \{t_i \mid t_i \in R \wedge t_i \in S\}$$

• 运行结果分析: 行、列?

A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁

A	B	C
a ₁	b ₂	c ₂
a ₁	b ₃	c ₂
a ₂	b ₂	c ₁

R.A	R.B	R.C	S.A	S.B	S.C
a ₁	b ₁	c ₁	a ₁	b ₂	c ₂
a ₁	b ₁	c ₁	a ₁	b ₃	c ₂
a ₁	b ₁	c ₁	a ₂	b ₂	c ₁
a ₁	b ₂	c ₂	a ₁	b ₂	c ₂
a ₁	b ₂	c ₂	a ₁	b ₃	c ₂
a ₁	b ₂	c ₂	a ₂	b ₂	c ₁
a ₂	b ₂	c ₁	a ₁	b ₂	c ₂
a ₂	b ₂	c ₁	a ₁	b ₃	c ₂
a ₂	b ₂	c ₁	a ₂	b ₂	c ₁

Hefei University of Technology

计算机与信息学院



2.4 关系代数语言

2.4.2 专门的关系运算

符号说明:

✓ R, t ∈ R, A = {A₁, A₂, ..., A_n}, t[A_i], t[A], t_i, t_j

✓ 象集Z_x

给定一个关系R (X, Z), X和Z为属性组。

当t[X] = x时, x在R中的象集 (Images Set) 为:

$$Z_x = \{t[Z] \mid t \in R, t[X] = x\}$$

它表示R中属性组X上值为x的诸元组在Z上分量的集合。

合肥工业大学

计算机与信息学院



2.4 关系代数语言

R	
x ₁	Z ₁
x ₁	Z ₂
x ₁	Z ₃
x ₂	Z ₂
x ₂	Z ₃
x ₃	Z ₁
x ₃	Z ₃

✓ x₁在R中的象集

$$Z_{x_1} = \{Z_1, Z_2, Z_3\},$$

✓ x₂在R中的象集

$$Z_{x_2} = \{Z_2, Z_3\},$$

✓ x₃在R中的象集

$$Z_{x_3} = \{Z_1, Z_3\}$$

合肥工业大学

计算机与信息学院



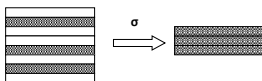
2.4 关系代数语言

(5) 选择 (σ) :

在关系R中选择满足给定条件的元组。

$$\sigma_F(R) = \{t \mid t \in R \wedge F(t) = \text{'真'}\}$$

F: 选择条件, 基本形式为: X₁ θ Y₁



合肥工业大学

计算机与信息学院



2.4 关系代数语言

[例1] 查询计算机专业 ('01') 全体学生。

$$\sigma_{\text{mno} = '01'}(\text{Students})$$

或 $\sigma_{S = '01'}(\text{Students})$

[例2] 查询软件工程专业全体男生信息。

$$\sigma_{\text{mno} = '03' \wedge \text{sex} = '男'}(\text{Students})$$

合肥工业大学

计算机与信息学院



2.4 关系代数语言

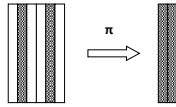
(6) 投影(π)

从 R 中选择出若干属性列组成新的关系

$$\pi_A(R) = \{ t[A] \mid t \in R \}$$

A : R 中的属性列

注意: 投影之后不仅取消了原关系中的某些列, 而且还可能取消某些元组 (避免重复行)



2.4 关系代数语言

[例3] 查询学生的姓名和所学专业编号。

$$\pi_{\text{name, mno}}(\text{Students})$$

或 $\pi_{2,5}(\text{Students})$

[例4] 查询学生关系 Students 中都有哪些专业。

$$\pi_{\text{mno}}(\text{Students})$$

(自动取消重复行!)

2.4 关系代数语言

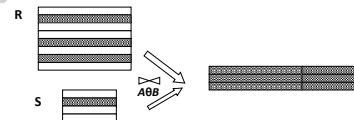
(7) 连接

- ✓ 连接也称为 θ 连接 (条件连接)
- ✓ 连接运算的含义: 从两个关系的笛卡尔积中选取属性间满足一定条件的元组。

$$R \bowtie_{A \theta B} S = \{ \langle t_r, t_s \rangle \mid t_r \in R \wedge t_s \in S \wedge t_r[A] \theta t_s[B] \}$$

- A 和 B : 分别为 R 和 S 上度数组等且可比的属性组
- θ : 比较运算符

2.4 关系代数语言



• 等值连接

$$R \bowtie_{A=B} S = \{ \langle t_r, t_s \rangle \mid t_r \in R \wedge t_s \in S \wedge t_r[A] = t_s[B] \}$$

• 自然连接: 一种特殊的等值连接!

$$R \bowtie S = \{ \langle t_r, t_s \rangle \mid t_r \in R \wedge t_s \in S \wedge t_r[B] = t_s[B] \}$$

2.4 关系代数语言

[例5] 设关系 R 和关系 S 如下:

R				S	
A	B	C	E	B	E
a_1	b_1	5	3	b_1	3
a_1	b_2	6	7	b_2	7
a_2	b_3	8	10	b_3	10
a_2	b_4	12	2	b_4	2

若改为等值
(自然)连接,
结果如何?

2.4 关系代数语言

➢ 外连接 \bowtie°

如果把舍弃的元组也保存在结果关系中, 而在其他属性上填空值(Null), 这种连接就叫做外连接。

✓ 左外连接 \bowtie°

如果只把左边关系 R 中要舍弃的元组保留就叫做左外连接。

✓ 右外连接 \bowtie°

如果只把右边关系 S 中要舍弃的元组保留就叫做右外连接。

2.4 关系代数语言

A	B	C	E
a ₁	b ₁	5	3
a ₁	b ₂	6	7
a ₂	b ₃	8	10
a ₂	b ₃	8	2

自然连接

A	B	C	E
a ₁	b ₁	5	3
a ₁	b ₂	6	7
a ₂	b ₃	8	10
a ₂	b ₃	8	2
NULL	b ₃	NULL	2

右外连接

合肥工业大学

计算机与信息学院

2.4 关系代数语言

(8) 除 (÷)

给定关系 $R(X, Y)$ 和 $S(Y, Z)$, 其中 X, Y, Z 为属性组。 R 与 S 中 Y 可以有不同属性名, 但必须出自相同的域集。 R 与 S 的除运算得到一个新的关系 $P(X)$, P 是 R 中满足下列条件的元组在 X 属性列上的投影:元组在 X 上分量值 x (x 的象集 Y_x 包含 S 在 Y 上投影) 的集合, 记作:

$$R \div S = \{ \pi_X[A] \mid \pi_Y \in R \wedge \pi_Y(S) \subseteq Y_x \}$$

 Y_x : x 在 R 中的象集, $x = \pi_X[x]$

合肥工业大学

计算机与信息学院

2.4 关系代数语言

设关系 R, S 分别为下图的(a)和(b), $R \div S$ 的结果为图(c)

A	B	C
a ₁	b ₁	c ₂
a ₂	b ₃	c ₇
a ₃	b ₄	c ₆
a ₁	b ₂	c ₃
a ₄	b ₆	c ₆
a ₂	b ₂	c ₃
a ₁	b ₂	c ₁

(a)

B	C	D
b ₁	c ₂	d ₁
b ₂	c ₁	d ₁
b ₂	c ₃	d ₂

(b)

R ÷ S
A
a ₁

(c)

合肥工业大学

计算机与信息学院

2.4 关系代数语言

✓ 在关系 R 中, A 可以取四个值 {a₁, a₂, a₃, a₄} a_1 的象集为 {(b₁, c₂), (b₂, c₃), (b₂, c₁)} a_2 的象集为 {(b₃, c₇), (b₂, c₃)} a_3 的象集为 {(b₄, c₆)} a_4 的象集为 {(b₆, c₆)}✓ S 在 (B, C) 上的投影为:{(b₁, c₂), (b₂, c₁), (b₂, c₃)}✓ 只有 a₁ 的象集包含了 S 在 (B, C) 属性组上的投影, 所以:

$$R \div S = \{ a_1 \}$$

合肥工业大学

计算机与信息学院

2.4 关系代数语言

聚合函数

count, sum, avg, min, max

$$\sigma_{\text{sum(Grade), max(Grade)}}(\text{SC})$$

分组聚集

$$\text{cno } \sigma_{\text{sum(Grade), max(Grade)}}(\text{SC})$$

合肥工业大学

计算机与信息学院

2.4 关系代数语言

数据修改操作

✓ 插入: $r \leftarrow r \cup E$ $\text{sc} \leftarrow \text{sc} \cup \{ \langle \text{"20090109"}, \text{"2006"}, \text{null} \rangle \}$ ✓ 删除: $r \leftarrow r - E$ $\text{students} \leftarrow \text{students} - \sigma_{\text{sno} = \text{"20090109"}}(\text{students})$ ✓ 更新: $r \leftarrow \pi_{F_1, F_2, \dots, F_n}(r), r \leftarrow \pi_{F_1, F_2, \dots, F_n}(\sigma_{p(r)}) \cup (r - \sigma_{p(r)})$ $\text{sc} \leftarrow \pi_{\text{sno}, \text{cno}, \text{Grade}+20}(\text{sc})$ $\text{sc} \leftarrow \pi_{\text{sno}, \text{cno}, \text{Grade}+10}(\sigma_{\text{Grade} \geq 50 \wedge \text{Grade} < 60}(\text{sc}))$ $\cup (\text{sc} - \sigma_{\text{Grade} \geq 50 \wedge \text{Grade} < 60}(\text{sc}))$

广义投影

合肥工业大学

计算机与信息学院



第2章 关系模型

■ 本章思考题：

从仓库管理中理解关系模型的完整性约束。

■ 本章作业：

P70 习题5、习题6（只要求用关系代数完成查询！）

